



KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0260

**BAYESIAN SEQUENTIAL *D-D* OPTIMAL
MODEL-ROBUST DESIGNS**

by

**A. RUGGOO
M. VANDEBROEK**

D/2002/2376/60

Bayesian sequential \mathcal{D} - \mathcal{D} optimal model-robust designs

Arvind Ruggoo

Martina Vandebroek

Department of Applied Economics, Katholieke Universiteit Leuven

Abstract

Alphabetic optimal design theory assumes that the model for which the optimal design is derived is usually known. However in real-life applications, this assumption may not be credible as models are rarely known in advance. Therefore, optimal designs derived under the classical approach may be the best design but for the wrong assumed model. In this paper, we extend Neff's (1996) Bayesian two-stage approach to design experiments for the general linear model when initial knowledge of the model is poor. A Bayesian optimality procedure that works well under model uncertainty is used in the first stage and the second stage design is then generated from an optimality procedure that incorporates the improved model knowledge from the first stage. In this way, a Bayesian \mathcal{D} - \mathcal{D} optimal model robust design is developed. Results show that the Bayesian \mathcal{D} - \mathcal{D} optimal design is superior in performance to the classical one-stage \mathcal{D} -optimal and the one-stage Bayesian \mathcal{D} -optimal design. We also investigate through a simulation study the ratio of sample sizes for the two stages and the minimum sample size desirable in the first stage.

Keywords: sequential designs, two-stage procedure, \mathcal{D} - \mathcal{D} optimality, posterior probabilities, model-robust.

1 Introduction

Many scientific investigations study processes in which the nature of the relationship between the process outcome and number of explanatory variables is unknown. In such situations, a collection of statistical techniques commonly referred to as Response Surface Methodology (RSM) has been developed (see Myers and Montgomery (1995) and Myers (1999) for more details).

More specifically, RSM is based on the premise that there exists some true physical relationship between the expectation $E(y) = \eta$ of the response of interest y as a function of controllable inputs \mathbf{x} , via physical constants $\boldsymbol{\theta}$

$$\eta = E(y) = h(\mathbf{x}, \boldsymbol{\theta}).$$

The response surface $h(\mathbf{x}, \boldsymbol{\theta})$ is typically unknown and may be in fact complex. Through a Taylor series expansion, it is assumed that within a region of interest the response surface η can be approximated by a linear graduating polynomial model of the form

$$\eta = E(y) \doteq \mathbf{f}'(\mathbf{x})\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a vector of model parameters to be estimated and \mathbf{f} is the polynomial expansion of the inputs \mathbf{x} . Under this modelling assumption a number of techniques has been developed for general exploration of the response surface, which provides invaluable knowledge in several applications in industry. The three major stages in RSM involve (1) data collection, which can be achieved by using carefully designed experiments (2) data analysis, which is related to model building and (3) optimization, having to do with searching for the best combination of input variables to optimize the response. In the sequel of this paper, we however concentrate on the data collection phase, where experimental design theory is involved.

An experimental design basically involves the specification of all aspects of an experiment. These may include the number of experimental units needed, levels of input or control variables to be used and choice of blocking factors. This is why the design of an experiment is as much an art as it is a science. The fundamental idea behind successful experimental design is that statistical inference about quantities of interest can be improved by appropriate choice of the input or control variables.

In all areas of experimental research work, costs and other limitations naturally dictates the parsimonious use of the available resources. Much of the statistical work on response surfaces in the literature has concerned the use of optimal design theory. The theory involves optimal selection of values of design factors under consideration within available resources.

Design optimality has gained momentum following the work by Kiefer (1958, 1959) and Kiefer and Wolfowitz (1959,1960) and has since become established in the statistical literature. To apply optimal design theory in practice necessitates a criterion for comparing experiments and an algorithm for optimising the criterion over the set of experimental designs.

In the context of the classical linear model theory, the statistical model that describes the assumed relationship between the response variable denoted as y , and the m factors or independent variables x_1, \dots, x_m is usually written as

$$y = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} + \varepsilon, \quad (1.1)$$

where $\mathbf{f}(\mathbf{x})$ is the $p \times 1$ vector representing the polynomial expansion of $\mathbf{x} = [x_1, \dots, x_m]'$ and $\boldsymbol{\beta}$ the $p \times 1$ vector of parameters $[\beta_1, \dots, \beta_p]'$. Because the observations are subject to random variation, the error term ε is added. The i th observation y_i can now be written as

$$y_i = \mathbf{f}'(\mathbf{x}_i)\boldsymbol{\beta} + \varepsilon_i, \quad (1.2)$$

where \mathbf{x}_i denotes the factor level combination for the i th observation. In experimental design, factor level combination \mathbf{x}_i is referred to as the design point associated with the i th observation. For n observations, it is convenient to rewrite (1.1) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1.3)$$

where $\mathbf{y} = [y_1, \dots, y_n]'$ denotes the $n \times 1$ vector of observations, \mathbf{X} is the extended design matrix of size $n \times p$ and $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_n]'$ is the $n \times 1$ vector of random error terms. The random error terms are usually assumed to be independent and identically distributed with zero mean and common variance σ^2 , i.e. $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and the variance-covariance matrix of the error terms is equal to $\text{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n$.

In the full rank model, $\mathbf{X}'\mathbf{X}$ is nonsingular and (1.3) has a unique solution, namely

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (1.4)$$

which is equivalent to the maximum likelihood estimator under normal errors.

The variance-covariance matrix of $\hat{\beta}$ is equal to

$$\text{cov}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (1.5)$$

Based on the parameter estimates (1.4), the predicted response at \mathbf{x} is given by

$$\hat{y}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\beta} \quad (1.6)$$

and a measure of the precision of the predicted value $\hat{y}(\mathbf{x})$, usually referred to as the prediction variance is

$$\text{var} \{ \hat{y}(\mathbf{x}) \} = \sigma^2 \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}). \quad (1.7)$$

Classical design optimality theory involves selecting the rows of \mathbf{X} so as to optimize some function of the Fisher's information matrix, $(\mathbf{X}'\mathbf{X})$.

With the advent of computer-generated designs in the 1980's, various algorithms have been written to make these optimal designs available for practitioners. However these computer generated designs have also suffered serious setbacks with the criticism that the criteria they use are heavily model dependent. One of the major arguments against the use of optimal design theory is the need to specify a model for the response function, coupled with the fact that optimal designs are frequently quite sensitive to the form of the model (See for example Box (1982) or the historical review by Myers, Khuri and Carter (1989)). Usually the knowledge of the form of \mathbf{f} in (1.1) is a weak one, since there may be little regressor knowledge before conducting the experiment. This makes it difficult to use optimality criteria in practice for RSM, as overspecification of \mathbf{f} may result in a design with some observations wasted trying to estimate unimportant parameter effect or in the opposite scenario, underspecification of \mathbf{f} may result in an inadequate model with not all necessary parameters being estimated.

It is thus fundamental that designs which are less dependent on the effect of model misspecification be developed. Box and Draper (1959) were the first authors to consider this issue in depth. They argued that a more appropriate criterion for such model-robust designs is one which uses the average mean squared error over the region of interest. We include some non-exhaustive literature on the topic for the interested reader: Laüter (1974) proposed using an average criterion that measures the precision of parameter estimation and also deals with effective discrimination for polynomial models. Huber (1975)

has investigated the effect of model misspecification by conducting a minimax analysis. He found that optimal designs based on first-degree polynomials may be subject to considerable bias from quadratic terms. Cook and Nachtsheim (1982) and Dette and Studden (1995) provided some analytical solutions to Läuter designs and expanded her work to include other robust criteria. Steinberg and Hunter (1984) give a nice overview with extensive references on these model-robust and model-sensitive designs.

The Bayesian approach to design optimality has gained popularity among research workers over the recent years as a way to address this problem of strong model dependence of optimality criteria. Covey-Crump and Silvey (1970) introduced the notion of Bayesian \mathcal{D} -optimality in the linear model case. O'Hagan (1978) postulated a Bayesian model in which the dependence of β on \mathbf{x} is characterised by a prior probability distribution that reflects beliefs about the likely smoothness and stability of the true response function. The work of Chaloner (1984) who derived Bayesian design optimality criteria for estimation and prediction for the linear model has led to progress within the Bayesian optimality paradigm over the years. An extensive bibliography of Bayesian designs is present in her work. Another excellent review is found in Chaloner and Verdinelli (1995) that explain the basis of Bayesian optimality with several practical applications. The basic idea to Bayesian design optimality is to choose the design that maximizes posterior information about some or all of the β 's conditional on the prior information available.

2 Overview of design optimality

In what follows, we shall first give an overview of non-Bayesian and Bayesian design optimality for the linear model and briefly review approaches for obtaining optimal designs when initial model knowledge is poor.

2.1 Non-Bayesian optimal design theory

Design optimality was introduced as early as 1918, by Smith. However, major contributions to the area are attributed to Kiefer (1958, 1959) and Kiefer and Wolfowitz (1959, 1960) who greatly extended previous work and introduced the alphabetic nomenclature for optimum design. Two approaches to optimal design theory are found in the

literature. The first approach ignores the integrality constraint of the number of observations at each design point. The resulting designs are called continuous or approximate. In the second approach, designs are constructed for a specified number of observations n . These designs are referred to as discrete or exact designs. In the sequel of this paper, we shall consider exact designs, since they are of practical relevance. However, knowledge of continuous designs is vital in offering insights into the construction of good practical designs for finite n .

Another important result in optimal design theory is the general equivalence theorem of Kiefer and Wolfowitz (1960) in which a design is represented by a probability measure on the predictor variable space (See Atkinson and Donev (1992) for further details).

Both (1.4) and (1.7) depend on the experimental design only through the $p \times p$ matrix $(\mathbf{X}'\mathbf{X})^{-1}$ and consequently a good experimental design will be one for which this matrix is 'small' in some sense. Various real-valued functions have been suggested as a measure of the 'smallness' of $(\mathbf{X}'\mathbf{X})^{-1}$. The most popular measure of design criterion is the \mathcal{D} -optimality criterion which minimizes the determinant of the variance-covariance matrix of the parameter estimates. The corresponding value of the \mathcal{D} -optimality criterion is

$$\mathcal{D} = |(\mathbf{X}'\mathbf{X})^{-1}|. \quad (2.1)$$

Under the assumption of independent normal errors with constant variance, the determinant of $(\mathbf{X}'\mathbf{X})^{-1}$ is inversely proportional to the square of the volume of the confidence region on the regression coefficients. The volume of the confidence region is relevant as it reflects how well the set of coefficients are estimated. A small $|(\mathbf{X}'\mathbf{X})|$ and hence large $|(\mathbf{X}'\mathbf{X})^{-1}|$ implies poor estimation of β .

The \mathcal{A} -optimal design deals with the individual variances of the regression coefficients as opposed to the \mathcal{D} -optimal criterion which utilises both the variances and covariances of the information matrix. Since the variances of the regression coefficients appear on the diagonals of $(\mathbf{X}'\mathbf{X})^{-1}$, the \mathcal{A} -optimal criteria minimizes the trace of $(\mathbf{X}'\mathbf{X})^{-1}$. The corresponding \mathcal{A} -criterion value equals

$$\mathcal{A} = \text{tr } (\mathbf{X}'\mathbf{X})^{-1}.$$

The \mathcal{G} -optimality criterion minimizes the maximum of the prediction variance over the design region \mathcal{X} . The \mathcal{G} -criterion value is given by

$$\max \text{var} \{\hat{\mathbf{y}}(\mathbf{x})\} = \max \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}). \quad (2.2)$$

The \mathcal{V} -optimal design minimizes the average prediction variance over the design region \mathcal{X} . This criterion is sometimes referred to as the \mathcal{Q} criterion also. The \mathcal{V} -criterion value is given by

$$\mathcal{V} = \int_{\mathcal{X}} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) \, dx.$$

Excellent reviews of these alphabetic optimality criteria described above and the many others are given in Ash and Hedayat (1978), Atkinson (1982), Myers, Khuri and Carter (1989) and Atkinson and Donev (1992).

2.2 Bayesian optimal design theory

As with many other areas in statistics, Bayesian designs have experienced rapid growth over the years. The Bayesian approach provides a framework where prior information and uncertainties regarding unknown quantities can be incorporated in the choice of a design.

Lindley (1972) presented a two-part decision theoretic approach to experimental design, which provides a unifying theory for most work in Bayesian experimental design today. Lindley's approach involves specification of a suitable utility function reflecting the purpose and costs of the experiment; the best design is selected to maximize expected utility (Clyde, 2001).

For the linear model (1.3), the analogue of the widely known non-Bayesian optimality criteria discussed in Section 2.1, such as the \mathcal{D} and \mathcal{A} -optimality and others can be given a decision-theoretic justification following Lindley's argument. For example, a utility function based on the Shannon information (1948) leads to Bayesian \mathcal{D} -optimality (See Bernardo, 1979 for details). In the case of a quadratic loss function, a Bayesian generalization of the \mathcal{A} -optimality design criterion is obtained. For more details on the various classes of Bayesian optimal designs and related discussions, we refer to Chaloner (1984), Verdinelli (1992), Chaloner and Verdinelli (1995), Verdinelli (2000) and Clyde (2001).

2.3 Model robust Bayesian designs

We now briefly look at designs within the Bayesian paradigm which are robust in some sense to model uncertainty. Dette (1993) developed Bayesian \mathcal{D} -optimal and model robust designs in linear regression models. The interesting work by DuMouchel and Jones (1994) illustrates a very elegant use of Bayesian methods to obtain designs which are more resistant to the biases caused by an incorrect model. They assume that there are two kinds of model terms, namely primary and potential terms which are high and low priority terms in the model respectively. By conceiving a prior distribution on the model coefficients that takes these primary and potential terms under consideration, they generate a Bayesian \mathcal{D} -optimal design by maximising the posterior information matrix. Similar type of work has been done by Andere-Rendon, Montgomery and Rollier (1997) for mixture models where they show that the performance of their Bayesian designs are superior to standard \mathcal{D} -optimal designs by producing smaller bias errors and improved coverage over the factor space.

Another strategy to develop robust designs found in the literature is the use of Bayesian sequential procedures. With this approach, it is possible to develop designs in two or more stages that leads to less dependence on model specification. The idea behind the sequential designs is intuitively appealing in the sense that the experimenter could have the opportunity to revise the design and model in the course of the experiment. Box and Lucas (1959) were the first authors to introduce the idea of sequential designs. Box, Hunter and Hunter (1978) also recommend the sequential assembly of smaller experiments over designing large comprehensive experiments, whenever this is feasible.

Throughout the optimal design literature, the sequential approach has been studied and developed within the non-linear framework. This is natural as the classical alphabetic optimality criterion requires prior knowledge of the model parameters, β due to the non-linearity of the problem. In this context, several two-stage designs have been developed in the literature. Abdelbasit and Plackett (1983) suggested a two-stage procedure to derive optimal designs for binary responses. Minkin (1987) generalises the idea of Abdelbasit and Plackett and proposes an improvement to their two-stage proposal. Myers, Myers, Carter and White (1996) proposed a two-stage procedure for the logistic regression that uses \mathcal{D} -optimality in the first stage followed by \mathcal{Q} -optimality in the second. Letsinger

(1995) also developed two-stage designs for the logistic regression model. Sitter and Wu (1995) developed two-stage designs for quantal response studies where they may be insufficient knowledge on a new therapeutic treatment or compound for the dose levels to be chosen properly.

Development of two-stage designs for linear models has been limited in the literature. Neff (1996) developed Bayesian two-stage designs under model uncertainty for mean estimation models. Montepiedra and Yeh (1998) developed a two-stage strategy for the construction of \mathcal{D} -optimal approximate designs for linear models. Lin, Myers and Ye (2000) developed Bayesian two-stage \mathcal{D} - \mathcal{D} optimal designs for mixture models.

3 Bayesian two-stage designs under model uncertainty

We now look in detail at the approach of Neff (1996) for developing Bayesian two-stage \mathcal{D} - \mathcal{D} optimal designs in the context of linear models. For the linear model $y = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta} + \varepsilon$, a Bayesian two-stage design approach makes it possible to efficiently design experiments when initial knowledge of \mathbf{f} is poor. This is accomplished by using Bayesian \mathcal{D} -optimality in the first stage following the approach of DuMouchel and Jones (1994) and then the second stage \mathcal{D} -optimal design is generated from the first stage by incorporating improved model knowledge from the first stage design.

3.1 Selection of the first stage design

Suppose that in a linear regression model such as (1.3), the regressors believed to be important in modeling the response are called primary terms and the uncertain regressors in the model are referred to as potential terms. The linear model (1.3) can thus be written as

$$\mathbf{y} = \mathbf{X}_{pri} \boldsymbol{\beta}_{pri} + \mathbf{X}_{pot} \boldsymbol{\beta}_{pot} + \boldsymbol{\varepsilon}, \quad (3.1)$$

where \mathbf{y} is a vector of $(n \times 1)$ responses distributed as $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, $\mathbf{X} = (\mathbf{X}_{pri}|\mathbf{X}_{pot})$ is a matrix of $(n \times (p + q))$, $\boldsymbol{\beta}_{pri}$ is a $p \times 1$ vector and $\boldsymbol{\beta}_{pot}$ is a $q \times 1$ vector attached to the primary and potential terms respectively. A first stage \mathcal{D} -optimal design can be obtained along the lines described by DuMouchel and Jones (1994). We describe their procedure,

but first look at a scaling convention and some distributional assumptions they propose.

A preliminary scaling and centering is recommended by DuMouchel and Jones to minimize the correlation between primary and potential terms and to permit the use of a standard prior distribution on the coefficients. The idea of the scaling convention is to make each potential term and all primary terms orthogonal to each other over the set of candidate points. This is achieved as follows in practice: Let $\mathbf{X} = (\mathbf{X}_{pri} | \mathbf{X}_{pot})$ represent the set of candidate points in model space. Regressing the potential terms on the primary terms results in $\boldsymbol{\alpha} = (\mathbf{X}_{pri}' \mathbf{X}_{pri})^{-1} \mathbf{X}_{pri}' \mathbf{X}_{pot}$, the alias matrix measuring how confounded \mathbf{X}_{pot} is with \mathbf{X}_{pri} . Replacing \mathbf{X}_{pot} by $\mathbf{Z} = \mathbf{R} / (\max\{\mathbf{R}\} - \min\{\mathbf{R}\})$ where $\mathbf{R} = \mathbf{X}_{pot} - \boldsymbol{\alpha} \mathbf{X}_{pri}$, essentially eliminates the aliasing over the candidates and this makes the choice of prior distributions easier.

Since the primary terms are likely to be active and no particular directions of their effects are assumed, the coefficients of the primary terms are specified to have a diffuse prior distribution - that is an arbitrary prior mean and prior variance tending to infinity. On the other hand, potential terms are unlikely to have huge effects, and it is proper to assume that they have a prior distribution with mean $\mathbf{0}$ and a finite variance. The assumption $N(\mathbf{0}, \tau^2 \sigma^2 \mathbf{I})$ is appropriate. The parameter τ^2 will determine the choice of designs, since it reflects the degree of uncertainty associated with potential terms relative to σ^2 . The larger the value of τ^2 , the stronger the belief in the potential terms. Under the assumption that primary and potential terms are uncorrelated (following the scaling convention), the joint prior distribution assigned to $\boldsymbol{\beta}_{pri}$ and $\boldsymbol{\beta}_{pot}$ is $N(\mathbf{0}, \sigma^2 \tau^2 \mathbf{K}^{-1})$ where \mathbf{K} is a $(p + q) \times (p + q)$ diagonal matrix, whose first p diagonal elements are equal to 0 and the remaining q diagonal elements are equal to 1.

Given the above prior distribution, the posterior distribution for the model parameters is also normal (see Pilz (1991)) and

$$p(\boldsymbol{\beta} | \mathbf{y}, \sigma^2) \sim N \left[\left(\mathbf{X}' \mathbf{X} + \frac{\mathbf{K}}{\tau^2} \right)^{-1} \mathbf{X}' \mathbf{y}, \sigma^2 \left(\mathbf{X}' \mathbf{X} + \frac{\mathbf{K}}{\tau^2} \right)^{-1} \right]. \quad (3.2)$$

DuMouchel and Jones's Bayesian analogue to the classical \mathcal{D} -optimal design in the case that prior information is available is the design that numerically maximizes the determinant of $(\mathbf{X}' \mathbf{X} + \mathbf{K} / \tau^2)$ with respect to \mathbf{X} . The resulting Bayesian \mathcal{D} -optimal design will

be one which will support estimation of both β_{pri} and β_{pot} but given the structure of the \mathbf{K} matrix, higher priority will be given to the primary terms. It can also easily be seen that as τ approaches ∞ , the Bayesian \mathcal{D} -optimal design becomes equivalent to the classical \mathcal{D} -optimal design with all $(p + q)$ terms treated equally. In the case $\tau = 0$, we end up with the \mathcal{D} -optimal design for the primary terms model only.

3.2 Analysis of first stage design

Let us consider the usual linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where it is assumed that $\mathbf{y}_i|\boldsymbol{\beta} \sim N(\mathbf{X}_i\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ for each stage i ($i = 1, 2$) with n_1 and n_2 observations for the first and second stage respectively. \mathbf{X} is the $(n_1 + n_2) \times (p + q)$ extended design matrix for the combined stages. The first stage Bayesian \mathcal{D} -optimal design is selected as described in Section 3.1 by minimising $|(\mathbf{X}_1'\mathbf{X}_1 + \mathbf{K}/\tau^2)^{-1}|$.

The parameter τ is unknown in the first stage and will depend on the belief the experimenter has on the potential terms. The larger the value τ , the more certain these terms are present in the full model. DuMouchel and Jones (1994) use the default value of $\tau = 1$ in their calculations. In the context of the two-stage procedure, Neff (1996) recommends a value of $\tau = 5$ in both the first and second stage for its ability to produce designs which are robust to model misspecification. Lin, Myers and Ye (2000) recommends a value of $\tau = 1$ for mixture models.

Before observing the first stage data, the experimenter has a model with $(p + q)$ regressors. The true relationship between the response and the input variables could be one of the following three types: (i) the model contains only the p primary terms (ii) the model contains all the p primary terms and q potential terms, i.e. the full model or (iii) the p primary terms and a subset of the q potential terms. Note that the three possibilities always include the p primary terms as these are the terms for which the experimenter has strong beliefs and hence should be present in the model. The total number of plausible models is thus $m = 2^q$. Consequently each candidate model M_i contains all primary terms and a subset q_i ($0 \leq q_i \leq q$) of potential terms.

The attractiveness of sequential designs becomes evident now. Once the data from the first stage has been collected, the information from the analysis can be used as prior infor-

mation to reduce model uncertainty in the next stage. Model knowledge can be updated by scoring each of the plausible candidate models using posterior probabilities indicating the likelihood that a particular candidate model is actually predicting the response adequately. These resulting scores or posterior probabilities can then be incorporated as weights to a second stage criterion.

Box and Meyer (1993) suggest a general way for calculating the posterior probabilities of different candidate models within the framework of fractionated screening experiments. Neff (1996) adapted their approach couched within the sequential Bayesian designs paradigm. We now describe her approach and adapt it to incorporate uncertainty in τ^2 .

The Bayesian approach to model identification is as follows (See for e.g. Box and Tiao (1968)). Let us label the set of all candidate models as M_0, M_1, \dots, M_m where M_0 corresponds to the primary terms model only. Each model M_i contains the parameters $\beta_i = \begin{bmatrix} \beta_{pri} \\ \beta_{pot(i)} \end{bmatrix}$ where β_{pri} contains all the p primary terms and $\beta_{pot(i)}$ contains a subset q_i of the q potential terms so that the sampling distribution of data \mathbf{y}_1 , given model M_i is described by the probability density $f(\mathbf{y}_1|M_i, \beta_i)$. The prior probability of model M_i is $p(M_i)$, and the prior density of β_i is $f(\beta_i|M_i)$. In the context of the first stage design, the predictive density of \mathbf{y}_1 given model M_i is given by the expression

$$f(\mathbf{y}_1|M_i) = \int_{\mathcal{B}_i} f(\mathbf{y}_1|M_i, \beta_i) f(\beta_i|M_i) d\beta_i, \quad (3.3)$$

where \mathcal{B}_i is the set of all possible values of β_i . Given the first stage data, the posterior probability of the model M_i given the data \mathbf{y}_1 is then

$$p(M_i|\mathbf{y}_1) = \frac{p(M_i)f(\mathbf{y}_1|M_i)}{\sum_{j=0}^m p(M_j)f(\mathbf{y}_1|M_j)}. \quad (3.4)$$

The posterior probabilities $p(M_i|\mathbf{y}_1)$ provide a basis for model identification and tentatively plausible models are identified by their large posterior probabilities. Now, each candidate model M_i contains the p primary terms and q_i potential terms ($0 \leq q_i \leq q$). Let π be the prior probability that any one of the potential terms is active in the true

model. Assuming that any one of the potential term being active is independent of beliefs about the other terms, the prior probability $p(M_i)$ of model M_i is given by

$$p(M_i) = \pi^{q_i}(1 - \pi)^{q - q_i}, \quad i = 1, 2, \dots, q. \quad (3.5)$$

The value of π should be chosen to represent the proportion of potential terms to be active. Since the experimenter may expect only a few of the potential terms to be really active, appropriate values of π would be in the range from 0 to 0.50. Box and Meyer (1993) found sensible results for $\pi = 0.25$. The analysis can be repeated for different values of π , say 0.1, 0.25, 0.5 but in general, the same factors will be identified as potentially active. Neff (1996) as well as Lin, Myers and Ye (2000) use $\pi=1/3$ in their computations.

Since we have assumed a normal linear model, the probability density of \mathbf{y}_1 given M_i is given by

$$f(\mathbf{y}_1|M_i, \boldsymbol{\beta}_i) \propto \sigma^{-n_1} \exp \left[-(\mathbf{y}_1 - \mathbf{X}_i \boldsymbol{\beta}_i)'(\mathbf{y}_1 - \mathbf{X}_i \boldsymbol{\beta}_i) / 2\sigma^2 \right].$$

Integrating (3.3) along the lines shown in Box and Meyer (1992) and assuming that $\boldsymbol{\beta}|\sigma^2, \tau^2 \sim N(\mathbf{0}, \sigma^2 \tau^2 \mathbf{K}^{-1})$ following the scaling convention described previously, we get

$$f(\mathbf{y}_1|M_i) \propto \tau^{-q_i} |\mathbf{T}_i + \mathbf{X}_i' \mathbf{X}_i|^{-1/2} \left(\mathbf{S}(\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i' \mathbf{T}_i \hat{\boldsymbol{\beta}}_i \right)^{-(n_1-1)/2}, \quad (3.6)$$

where \mathbf{X}_i is the first stage design in model M_i space and

$$\mathbf{T}_i = \frac{1}{\tau^2} \begin{bmatrix} \mathbf{0}_{p \times p} & \mathbf{0}_{p \times q_i} \\ \mathbf{0}_{q_i \times p} & \mathbf{I}_{q_i \times q_i} \end{bmatrix}$$

and

$$\hat{\boldsymbol{\beta}}_i = E(\boldsymbol{\beta}_i|\mathbf{y}_1), \text{ assuming model } M_i = (\mathbf{T}_i + \mathbf{X}_i' \mathbf{X}_i)^{-1} \mathbf{X}_i' \mathbf{y}_1$$

$$\mathbf{S}(\hat{\boldsymbol{\beta}}_i) = \text{Residual Sum of Squares for model } M_i = (\mathbf{y}_1 - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i)'(\mathbf{y}_1 - \mathbf{X}_i \hat{\boldsymbol{\beta}}_i).$$

The resulting posterior probability for model M_i becomes

$$p(M_i|\mathbf{y}_1) = C \pi^{q_i} (1 - \pi)^{q - q_i} \tau^{-q_i} |\mathbf{T}_i + \mathbf{X}_i' \mathbf{X}_i|^{-1/2} \left(\mathbf{S}(\hat{\boldsymbol{\beta}}_i) + \hat{\boldsymbol{\beta}}_i' \mathbf{T}_i \hat{\boldsymbol{\beta}}_i \right)^{-(n_1-1)/2}, \quad (3.7)$$

where C is the normalisation constant that forces all probabilities to sum to one.

In practice to avoid floating-point overflow, one actually computes

$$p(M_i|\mathbf{y}_1) = C \frac{p(M_i)f(\mathbf{y}_1|M_i)}{p(M_0)f(\mathbf{y}_1|M_0)} \quad (3.8)$$

for each model M_i , where M_0 is the model with primary terms only. Dividing by the constant $p(M_0)f(\mathbf{y}_1|M_0)$ does not change the final result as all the posterior probabilities are eventually scaled to unity, but this prevents getting large numbers as intermediate results (Box and Meyer, 1993).

Using (3.5) and (3.6), the expression (3.8) becomes

$$p(M_i|\mathbf{y}_1) = C \left(\frac{\pi}{(1-\pi)\tau} \right)^{q_i} \frac{|\mathbf{X}_0'\mathbf{X}_0|^{1/2}}{|\mathbf{T}_i + \mathbf{X}_i'\mathbf{X}_i|^{1/2}} \left(\frac{\mathbf{S}(\hat{\beta}_i) + \hat{\beta}_i'\mathbf{T}_i\hat{\beta}_i}{\mathbf{S}(\hat{\beta}_0)} \right)^{-(n_1-1)/2}. \quad (3.9)$$

Recall that in the first stage, it is assumed that all coefficients of potential terms are specified to have a $N(\mathbf{0}, \tau^2\sigma^2\mathbf{I})$ prior. The value of τ^2 is unknown and a guess value based on prior beliefs of the experimenter is used. Neff (1996) uses the same value of $\tau = 5$ in both her first and second stage designs. We propose to update the value of τ^2 to yield $\hat{\tau}^2$ for use in the second stage. We follow the approach of Box and Meyer (1992) for modelling and updating the uncertainty of τ^2 . Similar type of approach for mixture models has been proposed by Lin, Myers and Ye (2000). A convenient update of $\hat{\tau}^2$ is obtained as the mode of the posterior density of $p(\tau^2|\mathbf{y}_1)$. Setting the prior density of $p(\tau^2)$ to be locally uniform, i.e. substituting it with a constant, results in the posterior density $p(\tau^2|\mathbf{y}_1)$ to be given approximately by

$$\begin{aligned} p(\tau^2|\mathbf{y}_1) &\propto p(\tau^2)p(\mathbf{y}_1|\tau^2) \approx p(\mathbf{y}_1|\tau^2) \\ &= \sum_{i=0}^m p(M_i)p(\mathbf{y}_1|M_i, \tau^2) \\ &\propto \sum_{i=0}^m \left(\frac{\pi}{(1-\pi)\tau} \right)^{q_i} \frac{|\mathbf{X}_0'\mathbf{X}_0|^{1/2}}{|\mathbf{T}_i + \mathbf{X}_i'\mathbf{X}_i|^{1/2}} \left(\frac{\mathbf{S}(\hat{\beta}_i) + \hat{\beta}_i'\mathbf{T}_i\hat{\beta}_i}{\mathbf{S}(\hat{\beta}_0)} \right)^{-(n_1-1)/2}. \end{aligned} \quad (3.10)$$

Since $\hat{\tau}$ is the value most likely to occur in the posterior density of τ^2 given \mathbf{y}_1 , it is a more reasonable value to be used in all calculations after the first stage. In practice to compute $\hat{\tau}$, we repeat the posterior probabilities calculations for $\hat{\tau}$ over a fine grid of τ values and extract the value of τ that maximizes (3.10).

3.3 Selection of second stage Bayesian \mathcal{D} - \mathcal{D} optimal design

In the first stage it is assumed that the joint prior distribution of $\beta|\sigma^2, \tau^2$ is $N(\mathbf{0}, \sigma^2\tau^2\mathbf{K}^{-1})$ and the resulting first stage posterior distribution of $\beta|\sigma^2, \tau^2$ is $N(\mathbf{b}_1, \mathbf{V}_1)$ where $\mathbf{b}_1 = (\mathbf{X}'_1\mathbf{X}_1 + \mathbf{K}/\tau^2)^{-1}\mathbf{X}'_1\mathbf{y}_1$ and $\mathbf{V}_1 = \sigma^2(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{K}/\tau^2)^{-1}$. Let the second stage prior distribution of β be the first stage posterior. Assume that \mathbf{y}_2 , the n_2 second stage observations are also normal, namely, $\mathbf{y}_2|\beta, \sigma^2 \sim N(\mathbf{X}_2\beta, \sigma^2\mathbf{I})$, where \mathbf{X}_2 is the second stage design expanded to full model form. Then, the second stage posterior distribution of β is also normally distributed with mean $\mathbf{b}_2 = (\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'_2\mathbf{X}_2 + \mathbf{K}/\hat{\tau}^2)^{-1}\mathbf{X}'_1\mathbf{y}_1$ and variance $\mathbf{V}_2 = \sigma^2(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'_2\mathbf{X}_2 + \mathbf{K}/\hat{\tau}^2)^{-1}$. Note that τ^2 has been updated to $\hat{\tau}^2$ in the second stage.

Since β contains all $(p+q)$ parameters of the full model, a Bayesian \mathcal{D} - \mathcal{D} optimal design for the full model is found by choosing \mathbf{X}_2 so as to minimize $|(\mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'_2\mathbf{X}_2 + \mathbf{K}/\hat{\tau}^2)^{-1}|$. However, the full model is only one of the candidate models and in most cases not the most appropriate. Let us consider our subset models M_0, M_1, \dots, M_m as discussed previously, with each model M_i defined by its parameters β_i . The posterior variance of β_i is now

$$\mathbf{V}_{2(i)} = \sigma^2(\mathbf{X}'_{1(i)}\mathbf{X}_{1(i)} + \mathbf{X}'_{2(i)}\mathbf{X}_{2(i)} + \mathbf{K}_i/\hat{\tau}^2)^{-1},$$

where $\mathbf{X}_{1(i)}$ and $\mathbf{X}_{2(i)}$ are the first stage and second stage design matrices respectively, expanded to model space M_i and $\mathbf{K}_i = \begin{bmatrix} \mathbf{0}_{p \times p} & \mathbf{0}_{p \times q_i} \\ \mathbf{0}_{q_i \times p} & \mathbf{I}_{q_i \times q_i} \end{bmatrix}$ is the $(p+q_i) \times (p+q_i)$ matrix. A Bayesian \mathcal{D} - \mathcal{D} optimal design for model M_i is the set of design points \mathbf{X}_2 which minimizes $\mathcal{D}_i = |\mathbf{V}_{2(i)}|$. Since the posterior Box and Meyer probabilities computed from first stage data as described before reflect model importance, they can be incorporated as weights to average the \mathcal{D} criterion when the second stage is selected. This is performed by choosing the second stage design points so as to minimize \mathcal{D}_i for each model M_i . In practice, this is done by choosing the second stage design points $\mathbf{X}_2 \in \mathbf{X}$ so as to minimize

$$\sum_{M_i} \mathcal{D}_i p(M_i|\mathbf{y}_1).$$

4 The two-stage design procedure

The two-stage design procedure is best understood by contrasting it with the single stage design. A single stage design of size N implies that all the N design points are optimally selected and executed as a group. Analysis and statistical inference are then based on the resulting N observations. In contrast, a two-stage design is one in which the N observations are obtained by a series of two experiments run sequentially of each other. The first stage experiment is optimally designed using a certain optimality criterion, then conditional on the information provided from the first stage, the second stage design is chosen to create certain optimal criteria conditions in the combined design. Statistical inferences are then made based on all N observations obtained from the two-stages, as if it was completed in a single stage (Neff, 1996).

4.1 Effect of τ in the first stage design

Before looking at some examples of two-stage designs, we briefly examine the effect of τ on the selection of the first stage Bayesian \mathcal{D} -optimal design. Intuitively the parameter τ^2 will determine the choice of designs as it reflects the degree of uncertainty associated with potential terms relative to σ^2 . The larger the value of τ^2 , the stronger the belief in the potential terms.

Let us suppose that the full model under consideration by the experimenter comprises the following regressors:

$$\mathbf{x}^{(f)} = \{1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1^2, x_2^2, x_3^2\},$$

with $p = 4$ primary terms, $\{1, x_1, x_2, x_3\}$ and $q = 6$ potential terms, $\{x_1x_2, x_1x_3, x_2x_3, x_1^2, x_2^2, x_3^2\}$. Twelve runs are allocated to the first stage. The resulting designs for various values of τ are shown in Figure 1.

It can be seen that for $\tau = 0.1$, implying very little belief in the potential terms, the Bayesian \mathcal{D} -optimal design points spans at the extremes replicating four of the corner points. This is well-known with \mathcal{D} -optimality which usually pushes design points to extremes. It is not possible to estimate the quadratic effects in the potential terms. The same design points are selected for values of τ up to 0.4. For $\tau = 0.5$, the overall centroid

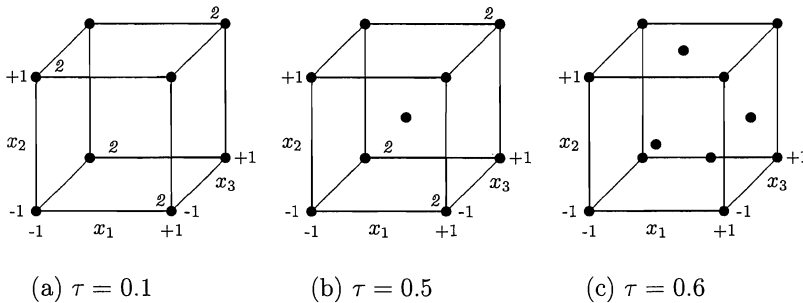


Figure 1: First stage designs with varying TAU values

is selected. The overall center point provides some information about the quadratic terms. However, it is still not possible to obtain independent information on the potential terms. When $\tau = 0.6$, three center points, one on each of the x_1 , x_2 , x_3 planes and a mid-edge point on the x_1 plane are selected. These additional four runs enable estimation of the potential terms. It could be seen that for $\tau > 0.6$, the eight corner points together with three center points and a mid-edge point on any of the planes are ever in the design. This abrupt change in the Bayesian \mathcal{D} -optimal solution at around $\tau = 0.6$ is also seen in the theoretical examples presented in Section 3 of the paper by DuMouchel and Jones (1994). It can easily be seen when $\tau < 0.4$, the Bayesian \mathcal{D} -optimal design will be identical to the classical \mathcal{D} -optimal for the primary terms model only. In case $\tau \geq 0.6$, the Bayesian \mathcal{D} -optimal design will become equivalent to the classical \mathcal{D} -optimal design with all $(p+q)$ terms treated equally, i.e. for the full model. The classical non-Bayesian \mathcal{D} -optimal designs obtained using PROC OPTEX in SAS for the primary terms and full model are shown in Figure 2 for comparison. The designs for $\tau = 0.1$ to 0.4 are identical to the primary terms model design as in Figure 2. With some rotation around the axes, the design for the full model in Figure 2 is similar to the Bayesian \mathcal{D} -optimal designs for $\tau \geq 0.6$.

We make the following concluding remark when an experiment is carried out solely in a one-stage procedure; the default value of $\tau = 1$ proposed by DuMouchel and Jones is a good compromise to use in that context. This is also supported by the work of Lin, Myers and Ye (2000) where the default value of $\tau = 1$ works well in the first stage design for mixture models. Neff (1996) recommends using $\tau = 5$ in a two-stage procedure. We shall investigate further the effect of τ on the combined two-stage design in Section 5.

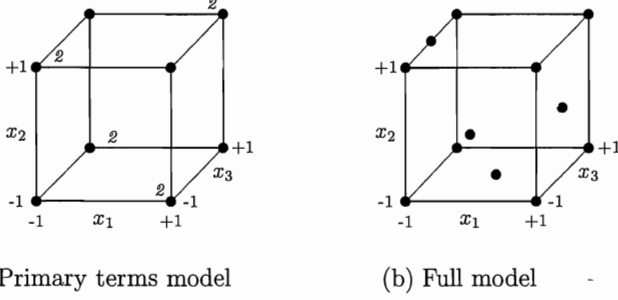


Figure 2: Non-Bayesian \mathcal{D} -optimal design for the primary terms and full model

4.2 Some examples of two-stage designs

We now present a few examples of Bayesian two-stage \mathcal{D} - \mathcal{D} optimal designs under different true models. Let us suppose that the full model under consideration by the experimenter comprises the following regressors:

$$\mathbf{x}^{(f)} = \{1, x_1, x_2, x_1^2, x_3, x_1x_2, x_1x_3, x_2x_3\}, \quad (4.1)$$

with $p = 4$ primary terms $\{1, x_1, x_2, x_1^2\}$ and $q = 4$ potential terms $\{x_3, x_1x_2, x_1x_3, x_2x_3\}$. Since the second stage design is a random variable depending on the first stage, response data from the first stage experiment are needed in the computation of the posterior probabilities used in the second stage criterion. We consider data simulated from three different true models under the assumption that $\varepsilon \sim N(0, 1)$. A moderate value of $\pi=1/3$ is used in all the examples.

CASE I

The first stage data set is simulated from the true model comprising only the primary terms $\{1, x_1, x_2, x_1^2\}$ namely,

$$y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 + \varepsilon.$$

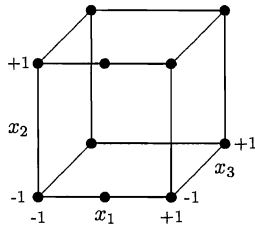


Figure 3: First stage design

The coefficients of the parameters of the true model are large compared to σ^2 (assumed to be unity) thereby indicating importance of the primary terms. Ten runs are allocated to each stage of the experiment and $\tau = 5$ is used in the first stage. The Bayesian optimality criterion presented in Section 3.1 is used to select the first ten runs. The resulting design is shown in Figure 3.

It can be seen that the design is desirable with all extreme points selected and two edge points on the x_1 plane for estimation of β_{11} in the full model (4.1). Only two levels of x_2 and x_3 are necessary since there is no interest in estimating β_{22} and β_{33} . According to Section 3.2, the response data from the first stage are used to get the value of $\hat{\tau}$. The results shown here are for one simulation data set only. Figure 4 shows the posterior density of τ^2 given \mathbf{y}_1 , which shows that the value of $\hat{\tau}$ with maximal probability is very close to zero. This is logical considering that the true model comprises primary terms only. The first stage data and value of $\hat{\tau}$ is used to compute the Box and Meyer posterior probabilities associated with each candidate models according to Section 3.2. The posterior probabilities are shown in Table 1 and we clearly see that the true model 'enjoys' the highest posterior probability. Several other simulations showed similar results.

For this simulation set, ten observations of a second stage design are chosen according to Section 3.3. Figure 5 shows the second stage design. Given the true model contains only primary terms, it can be seen that in the second stage design points are selected to better estimate these terms, namely several mid-edge points on the x_1 plane to estimate β_{11} adequately.

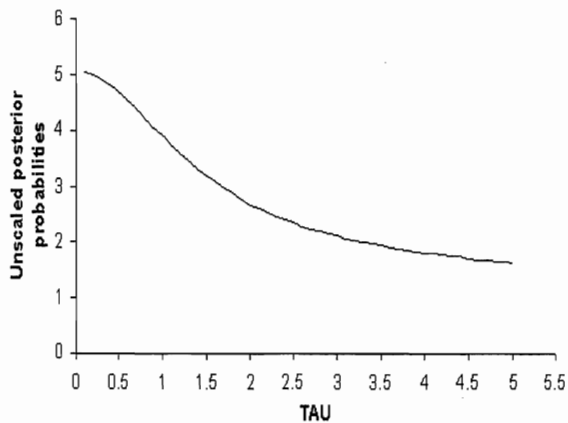


Figure 4: Posterior density of τ^2 given y_1 .

Table 1: Box and Meyer posterior probabilities for M_i

$p(M_i y_1)$	Terms in M_i
0.198	1 x_1 x_2 x_{11} (True model)
0.098	1 x_1 x_2 x_{11} x_3
0.098	1 x_1 x_2 x_{11} x_{12}
0.101	1 x_1 x_2 x_{11} x_{13}
0.098	1 x_1 x_2 x_{11} x_{23}
0.049	1 x_1 x_2 x_{11} x_3 x_{12}
0.050	1 x_1 x_2 x_{11} x_3 x_{13}
0.049	1 x_1 x_2 x_{11} x_3 x_{23}
0.050	1 x_1 x_2 x_{11} x_{12} x_{13}
0.049	1 x_1 x_2 x_{11} x_{12} x_{23}
0.050	1 x_1 x_2 x_{11} x_{13} x_{23}
0.025	1 x_1 x_2 x_{11} x_3 x_{12} x_{13}
0.024	1 x_1 x_2 x_{11} x_3 x_{12} x_{23}
0.025	1 x_1 x_2 x_{11} x_3 x_{13} x_{23}
0.025	1 x_1 x_2 x_{11} x_{12} x_{13} x_{23}
0.012	1 x_1 x_2 x_{11} x_3 x_{12} x_{13} x_{23}

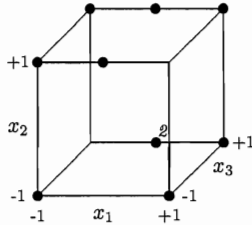


Figure 5: Second stage design

CASE II

The first stage data set is now simulated from the true model comprising all the primary terms and two potential terms $\{1, x_1, x_2, x_1^2, x_3, x_1x_2\}$ namely,

$$y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 1.80 x_3 + 1.3 x_1x_2 + \varepsilon.$$

The coefficients of the parameters of the potential terms x_3 and x_1x_2 are small compared to σ^2 and thus their effect will be only marginally significant in the true model. Again ten runs are allocated to each stage of the experiment and $\tau = 5$ in the first stage. The first stage design is shown as previously in Figure 3. According to Section 3.2, the response data from the first stage are used to get the value of $\hat{\tau}$ as previously. The results are for one simulation data set only. Figure 6 shows the posterior density of τ^2 given y_1 , which reveals that the value of $\hat{\tau}$ is very close to two. This makes sense considering that the true model comprises primary terms and two potential terms whose effects are only marginally significant compared to σ^2 given the parameters assigned to them in the true model. The first stage data and value of $\hat{\tau}$ are used to compute the measure of fit associated with each candidate models according to Section 3.2. The posterior probabilities are shown in Table 2 and we clearly see that the true model 'enjoys' the highest posterior probability. Several other simulations showed similar results.

For this simulation set, ten observations of a second stage design are chosen according to Section 3.3. Figure 7 shows the second stage design. The second stage design concentrates more points in estimating the terms in the true model, replicating one of the mid-edge point on the x_1 plane and selecting other mid-edge points on the x_1 plane so that esti-

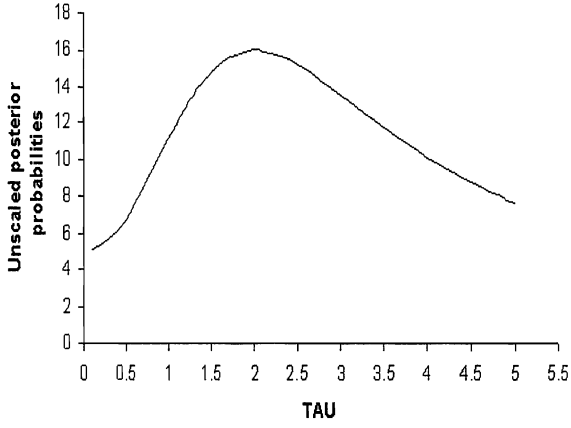


Figure 6: Posterior density of τ^2 given y_1 .

Table 2: Box and Meyer posterior probabilities for M_i

$p(M_i y_1)$	Terms in M_i
0.063	1 $x_1 x_2 x_{11}$
0.030	1 $x_1 x_2 x_{11} x_3$
0.159	1 $x_1 x_2 x_{11} x_{12}$
0.013	1 $x_1 x_2 x_{11} x_{13}$
0.015	1 $x_1 x_2 x_{11} x_{23}$
0.236	1 $x_1 x_2 x_{11} x_3 x_{12}$ (True model)
0.007	1 $x_1 x_2 x_{11} x_3 x_{13}$
0.008	1 $x_1 x_2 x_{11} x_3 x_{23}$
0.043	1 $x_1 x_2 x_{11} x_{12} x_{13}$
0.059	1 $x_1 x_2 x_{11} x_{12} x_{23}$
0.003	1 $x_1 x_2 x_{11} x_{13} x_{23}$
0.087	1 $x_1 x_2 x_{11} x_3 x_{12} x_{13}$
0.167	1 $x_1 x_2 x_{11} x_3 x_{12} x_{23}$
0.002	1 $x_1 x_2 x_{11} x_3 x_{13} x_{23}$
0.018	1 $x_1 x_2 x_{11} x_{12} x_{13} x_{23}$
0.089	1 $x_1 x_2 x_{11} x_3 x_{12} x_{13} x_{23}$

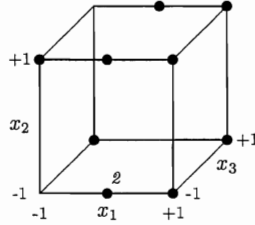


Figure 7: Second stage design

mation of β_{11} is very adequate as it has major importance with its large coefficient in the true model.

CASE III

The first stage data set is now simulated from the true model comprising the primary terms and the two potential terms as in Case II $\{1, x_1, x_2, x_1^2, x_3, x_1x_2\}$ namely,

$$y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 6.2 x_3 + 9.8 x_1x_2 + \varepsilon.$$

However, the coefficients of the parameters of the potential terms x_3 and x_1x_2 are now larger and thus their effect will be more significant in the true model as opposed to Case II where their effect were less prominent. Again ten runs are allocated to each stage of the experiment and $\tau = 5$ in the first stage. The response data from the first stage are used to get the value of $\hat{\tau}$ associated with each candidate model and the results shown are for one simulation data set. The posterior density of τ^2 given \mathbf{y}_1 (Figure 8) shows that the value of $\hat{\tau}$ is very close to twenty-two. This is natural considering that the true model comprises primary terms and two potential terms. However, the effects of the potential terms are now much more important compared to Case II. This obviously reflects very strong beliefs in these terms and consequently leads to the large value of $\hat{\tau}$ to be used in the second stage design. The Box and Meyer posterior probabilities associated with each candidate model are shown in Table 3 and we clearly see that the true model has the highest posterior probability.

The ten observations of the second stage design are shown in Figure 9. Since the potential

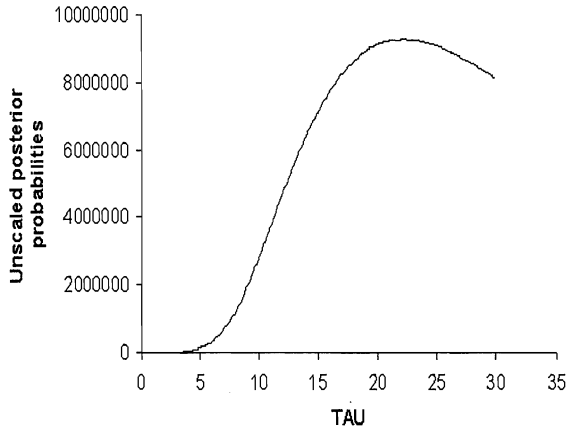


Figure 8: Posterior density of τ^2 given \mathbf{y}_1 .

Table 3: Box and Meyer posterior probabilities for M_i

$p(M_i \mathbf{y}_1)$	Terms in M_i
0	1 $x_1 x_2 x_{11}$
0	1 $x_1 x_2 x_{11} x_3$
0	1 $x_1 x_2 x_{11} x_{12}$
0	1 $x_1 x_2 x_{11} x_{13}$
0	1 $x_1 x_2 x_{11} x_{23}$
0.815	1 $x_1 x_2 x_{11} x_3 x_{12}$ (True model)
0	1 $x_1 x_2 x_{11} x_3 x_{13}$
0	1 $x_1 x_2 x_{11} x_3 x_{23}$
0	1 $x_1 x_2 x_{11} x_{12} x_{13}$
0	1 $x_1 x_2 x_{11} x_{12} x_{23}$
0	1 $x_1 x_2 x_{11} x_{13} x_{23}$
0.158	1 $x_1 x_2 x_{11} x_3 x_{12} x_{13}$
0.021	1 $x_1 x_2 x_{11} x_3 x_{12} x_{23}$
0	1 $x_1 x_2 x_{11} x_3 x_{13} x_{23}$
0	1 $x_1 x_2 x_{11} x_{12} x_{13} x_{23}$
0.006	1 $x_1 x_2 x_{11} x_3 x_{12} x_{13} x_{23}$

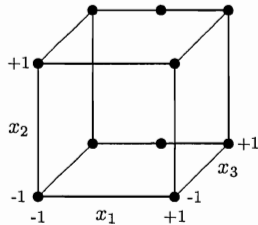


Figure 9: Second stage design

terms are more important than in Case II above, there is no replication of the mid-edge point on the x_1 plane as in Figure 7, but the design points in the second stage spans at the extremes again so that estimation of the potential terms parameters are adequate.

4.3 Conclusions from two-stage examples

The essence of all these examples is to demonstrate that the value of τ is updated accordingly in the second stage to reflect more belief in the potential terms, should they be important in the true model. In other words, $\hat{\tau}$ is small if the true model contains only primary terms and becomes larger as the marginal significance of the potential terms increases relative to σ^2 . Also we can deduce that in general, the second stage design is very effective in selecting design points based on information from the first stage design and considering terms present in the true model. The choice of τ in the combined design will be investigated in the next section.

Additional simulations were also conducted to study the influence of π on the two-stage procedure, by repeating the analyses above using $\pi = 0.1, 0.25, 0.3$ and 0.5 . In general, the procedure is quite robust to different values of π in the sense that usually the same design points are selected in the second stage. The nominal value of $\pi = 1/3$ has served well, but in any case experimenters should select π based on their prior expectations on the proportion of potential terms in the model.

5 Evaluation of Bayesian Two-Stage \mathcal{D} - \mathcal{D} Optimal Designs

The performance of the Bayesian two-stage \mathcal{D} - \mathcal{D} optimal designs presented in Sections 3.1 - 3.3 will now be evaluated relative to the classical one-stage designs. Since the second stage design is a random variable dependent on first stage data through the Box and Meyer posterior probabilities, we need a simulation approach to evaluate the performance of the two-stage procedure. We make use of the equality for the variance of a random variable \mathbf{Y} which is observed only in the presence of a random variable \mathcal{W}

$$\text{Var}(\mathbf{Y}) = E_{\mathcal{W}}[\text{Var}(\mathbf{Y}|\mathcal{W})] + \text{Var}_{\mathcal{W}}[E(\mathbf{Y}|\mathcal{W})].$$

In our context the variance expression becomes

$$\text{Var}(\hat{\beta}) = E_{\mathcal{W}}[\text{Var}(\hat{\beta}|\mathcal{W})] + \text{Var}_{\mathcal{W}}[E(\hat{\beta}|\mathcal{W})], \quad (5.1)$$

where \mathcal{W} is a vector comprising the second stage design points \mathbf{X}_2 and τ^2 , which are random variables dependent on first stage parameter estimates. Since $\hat{\beta}$ is the maximum likelihood estimate of β , equation (5.1) can be used in conjunction with a simulation approach to find the asymptotic variance-covariance properties of $\hat{\beta}$. Asymptotically, the second expression of (5.1) is zero as the expectation of $\hat{\beta}$ given \mathcal{W} is β . Therefore we have

$$\text{Var}(\hat{\beta}) \approx E_{\mathcal{W}}[\text{Var}(\hat{\beta}|\mathcal{W})]. \quad (5.2)$$

Hence $\text{Var}(\hat{\beta})$ for the two-stage procedure can be obtained by averaging $\text{Var}(\hat{\beta}|\mathcal{W})$ over numerous simulated two-stage designs.

The full model under consideration for the numerical evaluation is defined by the following set of regressors representing the full quadratic polynomial with 10 parameters.

$$\mathbf{x}^{(f)} = \{1, x_1, x_2, x_1^2, x_2^2, x_1x_2, x_3, x_1x_3, x_2x_3, x_3^2\}, \quad (5.3)$$

with $p = 5$ primary terms $\{1, x_1, x_2, x_1^2, x_2^2\}$ and $q = 5$ potential terms $\{x_1x_2, x_3, x_1x_3, x_2x_3, x_3^2\}$. We use $(p+q+2)$ as the sample size for the first stage as suggested by Neff and Myers (1998) for efficiency and robustness in the two-stage designs. For simulation

purposes we shall consider a full range of true models to evaluate the two-stage \mathcal{D} - \mathcal{D} procedure.

$$\text{Case I } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + \varepsilon.$$

$$\begin{aligned} \text{Case II } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & - 4.8x_3 + \varepsilon. \end{aligned}$$

$$\begin{aligned} \text{Case III } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & + 3.0x_1x_2 - 2.8x_3 + \varepsilon. \end{aligned}$$

$$\begin{aligned} \text{Case IV } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & + 1.8x_1x_2 - 1.4x_3 + 2.0x_1x_3 + \varepsilon. \end{aligned}$$

$$\begin{aligned} \text{Case V } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & + 9.6x_1x_2 - 8.4x_3 + 7.4x_3^2 + \varepsilon. \end{aligned}$$

$$\begin{aligned} \text{Case VI } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & + 3.8x_1x_2 - 1.4x_3 + 6.6x_3^2 + \varepsilon. \end{aligned}$$

$$\begin{aligned} \text{Case VII } y = & 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ & + 1.8x_1x_2 - 8.4x_3 + 2.0x_1x_3 + 5.6x_2x_3 - 7.4x_3^2 + \varepsilon. \end{aligned}$$

The true models above represent a wide range of possible model misspecifications. Since primary terms are believed to be important, they are assigned large coefficients compared to σ^2 (assumed to be unity). Potential terms coefficients are assigned small to moderate values to cover a large spectrum of possible model misspecifications. 100 data sets were simulated from each of the true models with each data set containing $n_1=12$ observations. Each of the simulations produced first stage data and the corresponding values of $\hat{\tau}$ and

consequently the posterior probabilities for use as the measure of fit in selecting the second stage design. We perform the evaluations with various values of τ 's in the first stage to reflect a range of level of uncertainty in the potential terms and also to investigate how the two-stage procedure performs with different τ values in the first stage following the discussion in Section 4.1. The values of τ 's used are 0.5, 0.6, 1.0, 5.0, 10.0 and $\pi = 1/3$ is used in all the computations. The error $\varepsilon \sim N(0, 1)$ is assumed in all the simulations. The size n_2 in the second stage design is chosen as 12 in the computations (Choice of n_1 and n_2 will be taken up in Section 6). The competitors to the Bayesian \mathcal{D} - \mathcal{D} optimal design are the one-stage traditional non-Bayesian \mathcal{D} -optimal design and one-stage Bayesian \mathcal{D} -optimal design designed optimally for the full model under consideration with design size $n = (n_1 + n_2) = 24$.

The performance of each design is measured by its design efficiency relative to the true models. The measure of efficiency is the usual \mathcal{D} -optimality criterion, namely

$$\mathcal{D}^* = |n(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}|.$$

where \mathbf{X}^* contains the $n = (n_1 + n_2)$ design points expanded to contain regressors only in the true model. Following (5.2), the performance of the two-stage procedures is measured by the average \mathcal{D}^* over the different 100 simulations, i.e.

$$\bar{\mathcal{D}}^* = \frac{\sum_{i=1}^{100} \mathcal{D}^*}{100}.$$

The standard errors of $\bar{\mathcal{D}}^*$ are also recorded. The one-stage traditional non-Bayesian \mathcal{D} -optimal design and one-stage Bayesian \mathcal{D} -optimal design are not data dependent and can thus be evaluated by a single \mathcal{D}^* , namely $\mathcal{D}^* = |n(\mathbf{X}^{*\prime}\mathbf{X}^*)^{-1}|$ for the true models.

Tables 4 - 10, show the results of the evaluation for the different true models defined above namely Cases I to VII. Values in brackets in the Tables 4 - 10, are the standard errors of $\bar{\mathcal{D}}^*$. In general the 100 simulations seemed adequate for assessing the performance of the two-stage designs as the standard errors were quite low in all the cases.

Since the Box and Meyer posterior probabilities are analogous to all subset regression in that all models being entertained are evaluated, choice of coefficients for the model

Table 4: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

Case I $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + \varepsilon$.

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	48.19 (0.217)*	57.44	51.84
0.6	48.37 (0.272)	57.44	51.84
1.0	47.95 (0.133)	57.44	51.97
5.0	47.82 (0.119)	57.44	57.44
10.0	47.95 (0.152)	57.44	57.44

* - Standard Error of $\bar{\mathcal{D}}^*$

parameters 'may' affect the sensitivity and stability of the simulations at times. However, for the range of models considered, results were quite consistent as seen with quite low standard errors in the simulations performed. For more insights on heuristics of instability and stabilization in model selection, see Breiman (1996).

The results from the tables show that the Bayesian \mathcal{D} - \mathcal{D} optimal design are more \mathcal{D} -efficient than the one-stage traditional non-Bayesian \mathcal{D} -optimal design and one-stage Bayesian \mathcal{D} -optimal design for all subsets of the true model, but not for the full model itself. However, even when the true model is the full model (Case VII), the two-stage procedure is very close in performance to its single stage competitors which were in fact constructed under the assumption that the full model is the true model. The procedures are quite robust to the different values of τ used in the first stage and work very well for both $\tau = 5$ and $\tau = 10$. We may thus recommend in the context of the two-stage

Table 5: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

Case II $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2$
 $- 4.8x_3 + \varepsilon.$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	56.06 (0.237)*	73.61	66.06
0.6	56.01 (0.252)	73.61	66.06
1.0	55.15 (0.248)	73.61	66.32
5.0	53.54 (0.268)	73.61	73.61
10.0	53.23 (0.236)	73.61	73.61

* - Standard Error of $\bar{\mathcal{D}}^*$

Table 6: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

Case III $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2$
 $+ 3.0x_1x_2 - 2.8x_3 + \varepsilon.$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	107.1 (0.62)*	119.7	116.0
0.6	107.5 (0.77)	119.7	116.0
1.0	106.6 (0.65)	119.7	115.9
5.0	104.8 (0.85)	119.7	119.7
10.0	104.7 (0.15)	119.7	119.7

* - Standard Error of $\bar{\mathcal{D}}^*$

Table 7: Values of \bar{D}^* for the two-stage design and single stage competitors

Case IV $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2$
 $+ 1.8x_1x_2 - 1.4x_3 + 2.0x_1x_3 + \varepsilon.$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	166.0 (1.58)*	195.2	189.4
0.6	165.7 (1.62)	195.2	189.4
1.0	163.8 (2.03)	195.2	189.3
5.0	155.1 (2.18)	195.2	195.2
10.0	157.4 (2.23)	195.2	195.2

* - Standard Error of \bar{D}^*

Table 8: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

$$\text{Case V } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 \\ + 9.6x_1x_2 - 8.4x_3 + 7.4x_3^2 + \varepsilon.$$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	650.7 (1.56)*	803.8	775.1
0.6	647.1 (1.67)	803.8	775.1
1.0	651.1 (1.58)	803.8	775.2
5.0	631.1 (1.70)	803.8	803.8
10.0	631.9 (1.74)	803.8	803.8

* - Standard Error of $\bar{\mathcal{D}}^*$

Table 9: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

Case VI $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2$
 $+ 3.8x_1x_2 - 1.4x_3 + 6.6x_3^2 + \varepsilon.$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	666.6 (3.67)*	803.8	775.1
0.6	663.4 (3.44)	803.8	775.1
1.0	665.3 (3.44)	803.8	775.2
5.0	643.1 (3.28)	803.8	803.8
10.0	640.5 (3.39)	803.8	803.8

* - Standard Error of $\bar{\mathcal{D}}^*$

Table 10: Values of $\bar{\mathcal{D}}^*$ for the two-stage design and single stage competitors

Case VII $y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2$
 $+ 1.8x_1x_2 - 8.4x_3 + 2.0x_1x_3 + 5.6x_2x_3 - 7.4x_3^2 + \varepsilon.$

TAU	Bayesian \mathcal{D} - \mathcal{D} procedure $n_1 = n_2 = 12$	One-Stage non-Bayesian \mathcal{D} $n = 24$	One-Stage Bayesian \mathcal{D} $n = 24$
0.5	2140 (10.5)*	2018	2077
0.6	2129 (9.1)	2018	2077
1.0	2142 (10.9)	2018	2077
5.0	2095 (5.8)	2018	2018
10.0	2114 (10.4)	2018	2018

* - Standard Error of $\bar{\mathcal{D}}^*$

procedure, the robust value of $\tau = 5$ in the first stage as proposed by Neff (1996) as it performs well in all simulations carried out. Also when the full model is chosen as the true model (Case VII), the \mathcal{D} -criterion value for $\tau = 5$ comes more closely to the values of the \mathcal{D} -criterion when either the one-stage non-Bayesian \mathcal{D} -optimal design or the one-stage Bayesian \mathcal{D} -optimal design are used.

We further make two very important remarks from our simulation studies about the choice of τ in the context of the one-stage and two-stage procedures:

1. In case the experiment is carried out solely in one-stage then use $\tau = 1$ as recommended by DuMouchel and Jones (1994). This is confirmed from our simulation studies where in Tables 4 - 10, the one-stage Bayesian \mathcal{D} -optimal design outperforms the classical \mathcal{D} -optimal design when $\tau = 1$ in all cases except when the true model is the full model. Use of $\tau = 5$ or $\tau = 10$ gave the same efficiency as the classical \mathcal{D} -optimal design and is therefore not recommended in a single stage procedure.
2. In the context of the two-stage procedures, if the experimenter wishes to use the Bayesian \mathcal{D} - \mathcal{D} optimality procedure and has no prior beliefs on the potential terms, then it is recommended to use the default value of $\tau = 5$ in the first stage to give robust designs to model misspecifications.

6 Ratio of Sample Size between two-stages

We now investigate in a simulation study the distribution of sample sizes between the two-stages and the number of design points desirable in the first stage. Letsinger (1995) and Myers et al. (1996) conclude from their study on logistic regression that the best performance of the two-stage designs was achieved when the first stage contained only 30% of the combined design size and 70% are reserved for the second stage. Neff (1996), without presenting the results of her simulations in her thesis recommends a 50% distribution of sample sizes in the stages to produce robust designs to model misspecifications. Lin, Myers and Ye (2000) also recommend a ratio of 1:1 between the two-stages to give satisfactory results for the mixture models.

Simulations are now used to investigate the reasonable ratio between the sample sizes and the minimum amount of observations required in the first stage. In all simulations that follow $\tau = 5$ in the first stage and consequently updated to $\hat{\tau}$ in the second stage, $\pi = 1/3$ and the error term ε is simulated from $N(0, 1)$.

It is assumed that the full model under consideration for the evaluations is defined by the following set of regressors representing a subset of the full quadratic polynomial comprising three input variables, namely

$$\mathbf{x}^{(f)} = \{1, x_1, x_2, x_1x_2, x_3, x_2x_3, x_1^2\}, \quad (6.1)$$

with $p = 4$ primary terms $\{1, x_1, x_2, x_1x_2\}$ and $q = 3$ potential terms $\{x_3, x_2x_3, x_1^2\}$. For simulation purposes we shall consider the following true models. Again primary terms are assigned large parameter values and potential terms moderate coefficient values compared to σ^2 , as they are less important.

$$\text{Case I} \quad y = 42.0 + 8.0 x_1 + 10.4 x_2 + 9.3 x_1x_2 + \varepsilon.$$

$$\text{Case II} \quad y = 42.0 + 8.0 x_1 + 10.4 x_2 + 9.3 x_1x_2 - 4.4 x_3 + \varepsilon.$$

$$\text{Case III} \quad y = 42.0 + 8.0 x_1 + 10.4 x_2 + 9.3 x_1x_2 - 4.4 x_3 + 5.8 x_1^2 + \varepsilon.$$

The total sample size of 20 available for the whole experiment is divided as follows and the five methods are considered below:

1. Bayesian \mathcal{D} - \mathcal{D} optimal design with $n_1=7$ and $n_2=13$.
2. Bayesian \mathcal{D} - \mathcal{D} optimal design with $n_1=9$ and $n_2=11$.
3. Bayesian \mathcal{D} - \mathcal{D} optimal design with $n_1=11$ and $n_2=9$.
4. Bayesian \mathcal{D} - \mathcal{D} optimal design with $n_1=13$ and $n_2=7$.
5. Bayesian \mathcal{D} - \mathcal{D} optimal design with $n_1=10$ and $n_2=10$.
6. Non-Bayesian one-stage \mathcal{D} -optimal design with $n=20$.

Table 11: Values of $\bar{\mathcal{D}}^*$ for the different designs

	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	\mathcal{D}
	procedure	procedure	procedure	procedure	procedure	procedure
	7-13	9-11	11-9	13-7	10-10	20
Case I	1.464	1.353	1.414	1.5625	1.305*	1.5625
Case II	1.522	1.371	1.417	1.5625	1.288*	1.5625
Case III	11.01	10.99	9.765*	9.765*	9.765*	9.765*

The best design is marked with an asterisk for each case

We use the notation $n_1 - n_2$ to represent the partition of the total sample size n between the two-stages. For e.g. a 10-12 partition would imply 10 runs in the first stage and 12 runs in the second. From Table 11, it can be seen that the 10-10 partition, implying a 1:1 ratio between the different stages works well compared to the other partitions. The values of $\bar{\mathcal{D}}^*$ for Case III are very close to that of the one-stage \mathcal{D} -optimal competitor, as we are considering six terms in the true model which is very near the full model with seven terms. Further, the 9-11 partition gives good results compared to other partitions, suggesting that a reasonable sample size for the first stage should be at least equal to $(p + q + 2)$ where p and q are the number of primary and potential terms respectively.

For the same models considered above, we now increase the total sample size to 30 and consider the following partitions 15-15, 12-18, 18-12, 14-16, 16-14. From Table 12, it can be seen that the 15-15 partition is most efficient for Cases I and II. The 16-14 performs better for Case III, though it is very close to the performance of the 15-15 partition.

We now examine with some additional simulations the larger full quadratic model with three input variables, namely

$$\mathbf{x}^{(f)} = \{1, x_1, x_2, x_1^2, x_2^2, x_1x_2, x_3, x_1x_3, x_2x_3, x_3^2\}, \quad (6.2)$$

with $p = 5$ primary terms $\{1, x_1, x_2, x_1^2, x_2^2\}$ and $q = 5$ potential terms $\{x_1x_2, x_3, x_1x_3, x_2x_3, x_3^2\}$. The total sample size available is 24 and this is partitioned as 12-12, 10-14 and 14-10. We consider the following five cases as true models for the simulation purposes. Again the primary terms are assigned large coefficient values relative to σ^2 and coefficients

Table 12: Values of $\bar{\mathcal{D}}^*$ for the different designs

	$\mathcal{D}-\mathcal{D}$ procedure	$\mathcal{D}-\mathcal{D}$ procedure	$\mathcal{D}-\mathcal{D}$ procedure	$\mathcal{D}-\mathcal{D}$ procedure	$\mathcal{D}-\mathcal{D}$ procedure	\mathcal{D} procedure
	15-15	12-18	18-12	14-16	16-14	30
Case I	1.3451*	1.3597	1.3494	1.3542	1.3484	1.8822
Case II	1.3576*	1.3619	1.3576*	1.3587	1.3586	1.9039
Case III	9.7420	9.7420	9.7420	9.7420	9.7418*	9.7769

The best design is marked with an asterisk for each case

of potential terms are assigned small to moderate values to reflect a range of possible model misspecification.

$$\text{Case I } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + \varepsilon.$$

$$\text{Case II } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 - 4.8x_3 + \varepsilon.$$

$$\text{Case III } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + 3.0x_1x_2 - 2.8x_3 + \varepsilon.$$

$$\text{Case IV } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + 9.6x_1x_2 - 8.4x_3 + 7.4x_3^2 + \varepsilon.$$

$$\text{Case V } y = 50.0 + 9.2 x_1 + 11.7 x_2 + 14.4 x_1^2 - 16.2 x_2^2 + 1.8x_1x_2 - 8.4x_3 + 2.0x_1x_3 + 5.6x_2x_3 - 7.4x_3^2 + \varepsilon.$$

From Table 13, the 12-12 partition is most efficient for Cases II and III. The 14-10 partition performs slightly better than the 12-12 partition for Case I and is the most efficient for Case IV. In case of the full model Case V, the one-stage procedure is most efficient, though the 12-12 partition is better than all other partitions in that case.

Table 13: Values of $\bar{\mathcal{D}}^*$ for the different designs

	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	$\mathcal{D}-\mathcal{D}$	\mathcal{D}
	procedure	procedure	procedure	procedure
	12-12	10-14	14-10	24
Case I	47.82	51.49	47.74*	57.44
Case II	53.54*	59.59	58.05	73.61
Case III	104.8*	116.8	107.6	119.7
Case IV	631.1	731.2	619.0*	803.8
Case V	2095	2527	2120	2018*

The best design is marked with an asterisk for each case

From the simulation studies described above and others conducted, the 1:1 partition between the two stages is a robust choice and a reasonable sample size for the first stage should be at least equal to $(p + q + 2)$.

We make a few additional observations based on our simulation results: too few runs in the first stage in general produce less efficient combined design, so if possible try to have at least $(p+q+2)$ in the first stage if resources permit. Trying to allocate more runs in the second stage than in the first generally leads to a decrease in efficiency of the combined design. Thus should practical situations prevent the use of an equal partition in the stages, we would recommend using a larger number of runs in the first stage than in the second stage. The reason is that allocating more resources to the first stage design, will generate a more efficient design robust to model misspecification. This in turn will give rise to better selection of the second stage design points via the Box and Meyer probabilities. The crucial idea in the two-stage procedure is thus to have adequate data collection in the first stage as the second stage is a random variable dependent on the first stage.

7 Discussion

Someone said, "The only time to design an experiment is after it has been run" as once the experimenter has an observation in hand, he begins to wish that he had collected the data differently. Therefore, designed experiments in general should be robust in na-

ture. This is exactly the contribution of this work in trying to achieve robust designs to model misspecifications using the Bayesian approach. This approach requires specification of prior distributions and parameters of those distributions and are recommended whenever prior knowledge can be expressed in a proper functional form. The choice of the various priors in our work are standard ones and work well for the two-stage approach.

In this paper, the two-stage Bayesian \mathcal{D} - \mathcal{D} strategy proposed by Neff (1996) for optimal designs has been extended by incorporating updated knowledge of τ^2 in the second stage based on first stage data. The Bayesian \mathcal{D} - \mathcal{D} optimal design developed satisfactorily deals with model misspecification as reflected by better performance in comparison to the classical one-stage \mathcal{D} -optimal procedures. We also found that the 1:1 partition is a reasonable ratio for the two-stage procedures and the first stage data should include a minimum of $(p+q+2)$ design points where p and q are the number of primary and potential terms respectively in the full model.

Multi-stage design are natural extensions of this procedure, however it is believed that these may not lead to increased improvements to justify practical use. Future work with the two-stage designs will involve incorporating other criteria to further reduce the dependency on the assumed model. These will include taking into account both model-robust and model-sensitive aspects in the design criterion by combining efficiency in estimating the primary terms, protection against bias caused by the potential terms and ability to test for lack-of-fit.

References

- Abdelbasit, K. M. and Plackett, R. L. (1983), "Experimental Designs for Binary Data," *Journal of the American Statistical Association*, 78, 90-98.
- Andere-Rendon, J., Montgomery, D.C. and Rollier, D.A. (1997), "Design of Mixture Experiments Usind Bayesian \mathcal{D} -Optimality," *Journal of Quality Technology*, 29, 451-463.
- Ash, A. and Hedayat, A. (1978), "An Introduction to Design Optimality with an Overview of the Literature, *Communications in Statistics - Theory and Methods*, 7, 1295-1325.
- Atkinson, A. and Donev, A. (1992). *Optimum Experimental Designs*. Clarendon Press, Oxford.
- Atkinson, A. C. (1982), "Developments in the Design of Experiments," *International Statistical Review*, 50, 161-177.
- Bernardo, J. M. (1979), "Expected Information as Expected Utility," *The Annals of Statistics*, 7, 686-690.
- Box, G. E. P. (1982), "Choice of Response Surface Design and Alphabetic Optimality," *Utilitas Mathematica*, 21, 11-55.
- Box, G. E. P. and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-653.
- Box, G. E. P., Hunter, W. G. and Hunter J. S. (1978). *Statistics for Experimenters*. John Wiley & Sons , New York.
- Box, G. E. P. and Lucas, H. L. (1959), "Design of Experiments in Non-Linear Situations," *Biometrika*, 46, 77-90.

- Box, G. E. P. and Meyer, R. D. (1993), "Finding the Active Factors in Fractionated Screening Experiments," *Journal of Quality Technology*, 25, 94-105.
- Box, G. E. P. and Meyer, R. D. (1992), "Finding the Active Factors in Fractionated Screening Experiments," *CQPI Report 80*, University of Wisconsin-Madison. Wisconsin.
- Box, G. E. P. and Tiao, G. C. (1968), "A Bayesian Approach to Some Outlier Problems," *Biometrika*, 55, 119-129.
- Breiman, L. (1996), "Heuristics of Instability and Stabilization in Model Selection," *The Annals of Statistics*, 24(6), 2350-2383.
- Chaloner, K. (1984), "Optimal Bayesian Experimental Design for Linear Models," *The Annals of Statistics*, 12, 283-300.
- Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 237-304.
- Clyde, M. A. (2001), "Experimental Design, A Bayesian Perspective," *ISDS Discussion Paper 01-05*, University of Duke.
- Cook, R. D. and Nachtsheim, C. (1982), "Model Robust, Linear-Optimal Designs," *Technometrics*, 24, 49-54.
- Covey-Crump, P. A. K. and Silvey, S. D. (1970), "Optimal Regression with Previous Observations," *Biometrika*, 57, 551-556.
- Detle, H. (1993), "Bayesian \mathcal{D} -Optimal and Model Robust Designs in Linear Regression Models," *Statistics*, 25, 27-46.
- Detle, H. and Studden, W. J. (1995), "Optimal Designs for Polynomial Regression when the Degree is not Known," *Statistica Sinica*, 5, 459-473.

- DuMouchel, W. and Jones, B. (1994), "A Simple Bayesian Modification of \mathcal{D} -Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics*, 36, 37-47.
- Huber, P. J. (1975), "Robustness and Designs," in *A Survey of Statistical Design and Linear Models* edited by N. Srivastava, North-Holland, New York, 287-301.
- Kiefer, J. (1958), "On the Nonrandomized Optimality and Randomized Non-optimality of Symmetrical designs," *Annals of Mathematical Statistics*, 29, 675-699.
- Kiefer, J. (1959), "Optimum Experimental Designs," *Journal of the Royal Statistical Society, Series B*, 21, 272-319.
- Kiefer, J. and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, 30, 271-294.
- Kiefer, J. and Wolfowitz, J. (1960), "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.
- Laüter, E. (1974), "Experimental Designs in a Class of Models," *Mathematische Operationsforschung und Statistik*, 5, 379-398.
- Letsinger, W. C. (1995). *Optimal One and Two-Stage Designs for the Logistic Regression Model*. Ph.D. dissertation, Virginia Polytechnic and State University, Virginia.
- Lin, H., Myers, R. H. and Ye, K. (2000), "Bayesian Two-Stage Optimal Design for Mixture Models," *Journal of Statistical Computation and Simulation*, 66, 209-231.
- Lindley, D. V. (1972), "Bayesian Statistics - A Review," *SIAM*, Philadelphia.
- Minkin, S. (1987), "Optimal Designs for Binary Data," *Journal of the American Statistical Association*, 82, 1098-1103.

Montepiedra, G. and Yeh, A. B. (1998), "A Two-Stage Strategy for the Construction of \mathcal{D} -Optimal Experimental Designs," *Communication in Statistics and Simulation*, 27(2), 377-401.

Myers, R. H. (1999), "Response Surface Methodology - Current Status and Future Directions (With Discussions)," *Journal of Quality Technology*, 31(1), 30-74.

Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology: Process and Product Improvement with Designed Experiments*. John Wiley & Sons, New York.

Myers, R. H., Khuri, A. I. and Carter, W. H. (1989), "Response Surface Methodology: 1966-1988," *Technometrics*, 31, 137-157.

Myers, W. R., Myers, R. H., Carter, W.H., Jr. and White, K. L, Jr. (1996), "Two-Stage Designs for the Logistic Regression Model in Single Agent Bioassays," *Journal of Biopharmaceutical Statistics*, 6(3), 283-301.

Neff, A. (1996). *Bayesian Two-Stage Designs under Model Uncertainty*. Ph.D. dissertation, Virginia Polytechnic and State University, Virginia.

Neff, A. and Myers, R. H. (1998). "Recent Developments in Response Surface Methodology and Its Application in Industry," in *Statistical Process Control for Quality and Productivity Improvement* edited by S. Park and G. G. Vining. Marcel Dekker. New York.

O'Hagan, A. (1978), "Curve Fitting and Optimal Design for Prediction," *Journal of the Royal Statistical Society, Series B*, 40, 1-41.

Pilz, J. (1991). *Bayesian Estimation and Experimental Designs*. John Wiley & Sons. New-York.

Shannon, C. E. (1948), "A Mathematical Theory of Communications," *Bell System Technical Journal*, 27, 379-423 and 623-656.

Sitter, R. R. and Wu, C. F. J. (1995), "Two-stage Design of Quantal Response Studies," *Technical Report 247*, University of Michigan, Michigan.

Smith, K. (1918), "On the standard deviation of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of observations," *Biometrika*, 12, 1-85.

Steinberg, D. and Hunter, W. (1984), "Experimental Design: Review and Comment," *Technometrics*, 26, 71-97.

Verdinelli, I. (1992), "Advances in Bayesian experimental design (with discussion)," in *Bayesian Statistics 4* edited by J. Bernardo, J. Berger, A. Dawid, A. Smith. Clarendon Press, Oxford.

Verdinelli, I. (2000), "A note on Bayesian design for the normal linear model with unknown variance," *Biometrika*, 87, 222-227.